

RESEARCH ARTICLE

Estimation of Esophageal Cancer Incidence in Tehran by Log-linear Method using Population-based Cancer Registry Data

Alireza Mosavi-Jarrahi¹, Toraj Ahmadi-Jouibari², Farid Najafi³, Yadollah Mehrabi¹, Abbas Aghaei^{2*}

Abstract

Background: Having knowledge or estimation of cancer incidence is necessary for planning and implementation of any cancer prevention and control programs. Population-based registries provide valuable information to achieve these objectives but require extra techniques to estimate the incidence rate. The present study aimed to estimate the esophageal cancer incidence using a log-linear method based on Tehran population-based cancer registry data. **Materials and Methods:** New cases of esophageal cancer reported by three sources of pathology reports, medical records, and death certificates to Tehran Metropolitan Area Cancer Registry Center during 2002-2006 were entered into the study and the incidence rate was estimated based on log-linear models. We used Akaike statistics to select the best-fit model. **Results:** During 2002-2006, 1,458 new cases of esophageal cancer were reported by the mentioned sources to the population-based cancer registry. Based on the reported cases, cancer incidence was 4.5 per 100,000 population and this was estimated to be 10.5 per 100,000 by the log-linear method. **Conclusions:** Based on the obtained results, it can be concluded that an estimated incidence for 2004 of 8.3 per 100,000 population could be a good benchmark for the incidence of esophageal cancer in the population of Tehran metropolis.

Keywords: Incidence estimation - log-linear - esophageal cancer - cancer registration - Akaike statistics

Asian Pac J Cancer Prev, 14 (9), 5367-5370

Introduction

Awareness on cancer incidence is essential like to plan, control, and promote regional and national cancer control programs (Kamo et al., 2007). An estimation of the incidence of all cases compared to registered cases or in other words, completeness level of cancer registration is one of the main parts of quality control in such registration. Incomplete record system faces the incidence and survival values calculation with distortion (Aghaei et al., 2012). This is important enough that frequently an important part of quality control in cancer registration reported by the International Association for Research on Cancer (IARC) is based on the estimated cases number in community- not based on the registered cases (Parkin, 1994; Schouten et al., 1994; Ghोजazadeh et al., 2013). Death certificate only and Mortality/Incidence Ratio indicators are among the indicators used normally to estimate the hidden population in registry systems (Shanmugaratnam, 1989), but they are indirect indicators that are strongly associated with the fatality rate of disease. When two or more incomplete lists of the number of cancer cases are available, capture-recapture (CR) method will be used to estimate the number of available cases in society, which is considered as a direct indicator for estimating cancer

incidence (Robles et al., 1988; Hilsenbeck et al., 1992; Laporte et al., 1992; Tull and McCarty, 1992; McCarty et al., 1993; Suwanrungruang et al., 2011). Among the techniques described in IARC reports, capture-recapture method has been considered at the same level of the best methods. During the last few decades, capture-recapture methods have been increasingly used in epidemiology and for the promotion of health care methods. Furthermore, they are used in various health fields to estimate hidden populations, incidence, and prevalence of diseases and special events. Such information is very important for policymakers and health workers (Hook and Regal, 1995; Aghaei et al., 2013). The use of capture-recapture methods is very efficient for reducing the costs of disease registration as well as reducing bias in incidence estimations and in the case of comparing population subgroups.

On the other hand, modeling the effect of intervening variables presents better estimations of the real size of population and therefore solves many problems of the estimation of population size (Tilling, 2001). Capture-recapture method has assumptions such as independence of sources, the equal chance of cases for being found in each source and independence of cases. Frequently, such assumptions are not expected for available sources, so log-

¹Department of Epidemiology, School of Public Health, Shaheed Beheshti University of Medical Sciences, Tehran, ²Clinical Research Development Center, Imam Khomeini Hospital, ³Department of Epidemiology, School of Public Health, Kermanshah University of Medical Sciences, Kermanshah, Iran *For correspondence: aqaei.a@gmail.com

linear models are proposed to be used in such situations (Laporte, 1994; Héraud-Bousquet et al., 2012).

The activity of Tehran's population-based cancer registry center or Tehran Metropolitan Area Cancer Registry (TMACR) has began since 1998. The main goal of this center is to measure cancer incidence in residents of Tehran metropolis. Data collection area of TMACR includes Tehran metropolis plus Islamshahr and Shemiranat townships (Etemadi et al., 2008; Aghaei et al., 2013).

Central areas of Iran including Tehran have the highest incidence of esophageal cancer in Iran after North and Northeast regions (Ramdar, 2010; Sadjadi et al., 2010). According to the first Tehran's population-based cancer registry report during 1998-2001, men and women living in areas covered by this center have allocated the sixth and the fifth ranks of malignant esophageal cancers to themselves, respectively.

This study is among the first researches done by log-linear method in the country and aims to estimate the incidence of esophageal cancer in TMACR.

Materials and Methods

New cases of esophageal cancer reported by three sources of death certificates, pathology reports, and medical records to TMACR during 2002-2006 entered the study. Since Tehran is considered as a referral center, cases not being among the areas covered by the registration center according to their address and contact number were excluded from the list. Duplicates were identified and removed using EXCEL software. Features such as name, surname, and father's name were used to identify common cases among sources and, if necessary, date of birth and ID number options were also used.

In this study, all three sources entered the Log-linear models. Regarding the possible relationships between resources in three-source method, there will be eight models. A model in which all three sources are independent; a model in which a pair of sources are interdependent and independent of the third source, of which there will be three modes; a model in which two pairs of sources are interdependent and independent of the third pair, of which there will be three modes too; and the eighth model is the case in which all binary relations will be. To select the best-fit model, we used Akaike statistics, the use of which is very common in scientific researches and analyses. (Hook and Regal, 1995; Hook and Regal, 1997).

The formula for this statistics in log-linear method is as follows: $AIC = -2G^2 + 2(df)$

Where G^2 is the statistical ratio of likelihood function related to models fitness and df is the degrees of model freedom. To select the model, we can act as the following. The model with the highest value of Akaike statistics and the highest p value for G^2 statistic is chosen as the best model. SPSS software version 16 was used to select the model and estimate the incidence.

Incidence is obtained based on the new cases of an event (esophageal cancer) occurred at a certain time over the number of the population at risk for same event at that

time. In this study, all the incidences are reported based on the incidence per one hundred thousand populations.

Then the incidence estimation was conducted based on the selected model for the subgroup of study calendar years, gender subgroup, and age subgroups. Denominators of the incidence for calendar years considered the population of that year and for gender and age subgroups are considered the population of five years of study.

Results

After investigating common cases between sources and removing duplicates, new cases of esophageal cancer reported to TMACR during 2002-2006 were equal to 1458. The average age of all cases was 70.43 years, 70.54 (± 11.47) years for men and 70.33 (± 12.4) years for women. The men to women ratio were 1.36:1.

A model in which two sources of pathology reports and medical records and also death certificates and medical records were mutually interdependent and two sources of pathology reports and death certificates were independent was selected as the best model with the highest value of

Table 1. Model Selection by G^2 Statistic and AIC Statistic

Model [†]	Estimate of incidence rate	G^2 statistic	p value for G^2 statistic	AIC statistic
*1/2/3	5.9	388.2	0.0001	-770.4
12/3	4.4	224.7	0.0001	-445.4
13/2	5.4	350.7	0.0001	-697.4
23/1	11.3	3.2	0.364	-2.4
12/23	11	3.2	0.206	-4.4
13/12	4	109	0.0001	-216
13/23	10.5	0.43	0.806	1.14
13/23/12	14.5	0	1	0

*Source 1: death certificates. Source 2: Pathology reports. Source 3: Medical records; [†]Model 1/2/3: A model where all available resources are independent; Model 12/3: A model where sources 1 and 2 are dependent and independent of the source 3; Model 13/2: A model where sources 1 and 3 are dependent and independent of the source 2; Model 23/1: A model where sources 2 and 3 are dependent and independent of the source 1; Model 12/23: A model where two sources 1 and 2 and also two sources 2 and 3 are mutually interdependent and two sources 1 and 3 are independent; Model 13/12: A model where two sources 1 and 3 and also two sources 1 and 2 are mutually interdependent and two sources 3 and 2 are independent; Model 13/23: A model where two sources 1 and 3 and also two sources 2 and 3 are mutually interdependent and two sources 2 and 1 are independent; Model 13/23/12: A model where all binary relations between resources are exist

Table 2. Estimating the Esophageal Cancer Incidence by Log-linear Method Based on the Population of Tehran Metropolis During 2002-2006

Subgroups	*Reported incidence rate	Estimated incidence rate (per 100,000)	95%CI for estimated incidence rate
Calendar year	2002	4.3	11.5 (8.2-14.8)
	2003	3.8	11.4 (7.8-15.0)
	2004	4.8	8.3 (7.0-9.6)
	2005	4.6	10.3 (7.4-13.2)
	2006	4.8	11.1 (7.6-14.6)
sex	Male	4.9	†13.9 (11.6-16.2)
	Female	4	7.9 (6.6-9.2)
Age group (year)	50 \geq	0.8	2.1 (1.3-2.9)
	51-64	10.4	23.8 (17.0-30.6)
	65 \leq	39.4	97.6 (75.0-120.2)
Hole		4.5	10.5 (8.6-12.4)

*Incidence reported by death certificates, pathology reports and medical records; [†]Incidence for gender and age subgroups is obtained in terms of the number of the population at risk during 5 years studied

Akaike statistics and the highest p-value for G^2 statistic (Table 1).

The incidence of esophageal cancer was estimated generally and to the studied subgroups based on the population estimation of 2002-2005 and the Census of Tehran metropolis in 2006 (Tehran, Islamshahr and shemiranat townships) (Table 2).

Discussion

There are two important and relevant assumptions when using capture-recapture methods in epidemiological studies. First, if two sources are used, they should be independent of each other and generally, in the case of using a model with K source, there should be no dependence between all K sources. Second, all people should have equal chance entering the resources. Underestimation or over-estimation proportional to the actual population size will be possible in the absence of each of these assumptions. These assumptions are very low likely to be fully established in an epidemiological study. For example, death certificates and hospital medical records documentation are two sources commonly used in a cancer registry system. Cases registered in each of the sources are very likely to be recorded in the other source. This shows dependency between the sources and flaw-making in the first assumption. On the other hand, the possibility of hospitalization of patients with severe disease in the hospital and their registration as cancer cases is much greater and also death because of disease will be more in these people, which is in conflict with the second assumption. These problems decrease accuracy and reliability of the estimation provided by capture-recapture method. Approach to encounter these problems is the use of more than two sources, the use of log-linear models for modeling dependencies between resources (Hook and Regal, 1995), and the use of layering methods in terms of disease behaviors and people for equalization of cases registration chance in each source (McClish and Penberthy, 2004).

The average age of esophageal cancer incidence was 70.43 (70.54 for men and 70.33 for women). The age of incidence does not show much difference between men and women, but is slightly higher than the average age reported by other studies conducted in other parts of the world (Bashash et al., 2008). The ratio of incidence in men to women is larger than one consistent with international and regional reports (Boyle and Levin, 2008; Babaei et al., 2009; Mohagheghi et al., 2009; Sadjadi et al., 2010; Amani et al., 2013; Wang et al., 2013).

The result of log-linear models fitness was as follows: on the one hand, the sources of medical records and pathology reports and on the other hand, medical records and death certificates are mutually interdependent and two sources of pathology reports and death certificates are considered independent. Although in this study, we have tried to select the best model using AIC index and p value for G^2 statistic, depiction of these relationships in terms of what happens in society seems also logical because it is very difficult to consider pathologies and hospitals independent. In addition, in death certificates

list, the possibility of finding a case with acute conditions having been hospitalized is much more than finding cases with better conditions.

According to WHO (World Health Organization) report in 2008, Blacks of America, the countries of South Africa and East and Central Asia have had the highest incidence. In addition, according to GLOBOCAN 2008 estimations providing an estimation of incidence and mortality related to cancer, the highest estimated incidence has been recorded to be for the countries of South Africa, East Asia and East Europe and the lowest incidence has been for West, Center and North Africa countries as well as countries located in Central America (Boyle and Levin, 2008). In this study, the registered incidence was 4.5 per one hundred thousand populations being less than the value estimated by WHO for Iran (6.8 per one hundred thousand populations) and also less than the values expected based on studies conducted in the country. Sajjadi et al. (2010) reported that northern regions of Iran have a high rate of esophageal cancer incidence, so that Golestan province (Iran) located in the Northern East and Ardabil province in the North West have the incidence of about 40 and 15 per one hundred thousand populations, respectively. Moving towards the southern parts of the country, esophageal cancer incidence is reduced, so that Semnan province in the center and Kerman province in the south of the country have the incidence of about 11 and 3 per one hundred thousand populations, respectively. Since Tehran metropolis is located in the center of the country and has placed a combination of all the country populations in itself because of its geographical, political and economic position, it is expected to have a medium incidence rate compared to other parts of the country and be the same as the central provinces of the country. While the incidence reported in this study and earlier TMACR reports is less than the expected values in the country's center, the estimated incidence by log-linear method is similar to the incidence of cancer in the country's central provinces (Mohagheghi et al., 2009; Sadjadi et al., 2010).

Incidence based on the studied calendar years are almost similar to each other, and have an increasing but irregular trend. The values estimated by log-linear method are higher than the values reported to the registration system indicating the low completeness of cancer registration in TMACR. The highest and lowest incidence was related to calendar years 2004 and 2002, respectively. Since death certificate is one of the three sources used in this study and 5-year survival rate of esophageal cancer patients is about 10% based on the studies (Bashash et al., 2008; Boyle and Levin, 2008), it is possible - in the first years of this study - that some cases which have been identified by death certificates to have been registered by other sources in the years before this study. On the other hand, the cases reported in the last years of this study by two sources of pathology reports and medical records are less likely to be observed in death certificates. Therefore, it is expected that the middle years of this study show more complete and consequently a more precise estimation.

The estimation of esophageal cancer incidence shows a salient difference between men and women. Since the estimated incidence in women is more similar to the

reported incidence, it is indicated that the registration completeness for women is larger than that for men. On the other hand, according to GLOBOCAN 2008 estimations the incidence of 6.3 and 7.4 per one hundred thousand populations for Iranian women and Iranian men reported, respectively. The estimation of this study for women is much similar to that estimation.

The reported incidence in the age subgroups is almost similar to the previous TMPCR reports (Mohagheghi et al., 2009). However the estimated incidence for each age group in present report is more than twice the reported values. In populations with very high incidence, the incidence report in older people is about 3-4 times the value estimated in this study (He et al., 2008; Matsuda et al., 2008). As expected, the estimated incidence for age subgroups is less than the reported values for northern areas of country that having high esophageal cancer incidence (Semnani et al., 2006; Babaei et al., 2009).

Based on what we have mentioned before about the precision of estimated incidence for middle year, we believe that the estimation obtained for calendar year of 2004 can be considered as a good estimation for esophageal cancer incidence in the area under the coverage of TMPCR.

References

- Aghaei A, Najaf F, Mosavi-Jarrahi A, Ahmadi-Jouibari T (2012). Coverage, density and completeness of sources used in tehran metropolitan area cancer registry: according to the data of esophageal cancer, 2003-2007. *Asian Pac J Cancer Prev*, **13**, 3617-9.
- Aghaei A, Ahmadi-Jouibari T, Baiki O, Mosavi-Jarrahi A (2013). Estimation of the gastric cancer incidence in Tehran by two-source capture-recapture. *Asian Pac J Cancer Prev*, **14**, 673-7.
- Amani F, Ahari SS, Akhghari L (2013). Epidemiology of esophageal cancer in ardebil province during 2003-2011. *Asian Pac J Cancer Prev*, **14**, 4177-80.
- Babaei M, Jaafarzadeh H, Sadjadi AR, et al (2009). Cancer incidence and mortality in Ardabil: report of an ongoing population-based cancer registry in Iran, 2004-2006. *Iranian J Publ Health*, **38**, 35-48.
- Bashash M, Shah A, Hislop G, et al (2008). Incidence and survival for gastric and esophageal cancer diagnosed in British Columbia, 1990 to 1999. *Can J Gastroenterol*, **22**, 143-8.
- Boyle P, Levin B (2008). International Agency for Research on, World cancer report. IARC press, 338.
- Etemadi A, Sadjadi A, Semnani S, et al (2008). Cancer registry in Iran: a Brief Overview. *Arch Iran Me*, **11**, 577-80.
- Ghojzadeh M, Mohammadi M, Azami-Aghdash S, et al (2013). Estimation of cancer cases using capture-recapture method in Northwest Iran. *Asian Pac J Cancer Prev*, **14**, 3237-41.
- He YT, Hou J, Chen ZF, et al (2008). Trends in incidence of esophageal and gastric cardia cancer in high-risk areas in China. *Eur J Cancer Prev*, **17**, 71-6.
- Héraud-Bousquet V, Lot F, Esvan M, et al (2012). A three-source capture-recapture estimate of the number of new HIV diagnoses in children in France from 2003-2006 with multiple imputation of a variable of heterogeneous catchability. *BMC Infect Dis*, **12**, 51.
- Hilsenbeck SG, Kurucz, C, Duncan RC (1992). Estimation of completeness and adjustment of age-specific and age-standardized incidence rates. *Biometrics*, **48**, 1249-62.
- Hook EB, Regal RR (1995). Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev*, **17**, 243-64.
- Hook EB, Regal RR (1997). Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation. *Am J Epidemiol*, **145**, 1138-44.
- Kamo K, Kaneko S, Satoh K, et al (2007). A mathematical estimation of true cancer incidence using data from population-based cancer registries. *Jpn J Clin Oncol*, **37**, 150-5.
- LaPorte RE, McCarty DJ, Tull ES, Tajima N (1992). Counting birds, bees, and NCDs. *Lancet*, **339**, 494-5.
- LaPorte RE (1994). Assessing the human condition: capture-recapture techniques. *BMJ*, **308**, 5-6.
- Matsuda T, Marugame T, Kamo K, et al (2008). Cancer incidence and incidence rates in Japan in 2002: based on data from 11 population-based cancer registries. *Jpn J Clin Oncol*, **38**, 641-8.
- McCarty DJ, Tull ES, Moy CS, Kwok CK, LaPorte RE (1993). Ascertainment corrected rates: applications of capture-recapture methods. *Int J Epidemiol*, **22**, 559-65.
- McClish D, Penberthy L (2004). Using multivariate capture-recapture techniques and statewide hospital discharge data to assess the validity of a cancer registry for epidemiologic use. *Health Serv Outcomes Res Methodol*, **5**, 141-52.
- Mohagheghi MA, Mosavi-Jarrahi A, Malekzadeh R, Parkin M (2009). Cancer incidence in Tehran metropolis: the first report from the Tehran population-based cancer registry, 1998-2001. *Arch Iran Med*, **12**, 15-23.
- Parkin DM (1994). Comparability and quality control in cancer registration: international agency for Research on Cancer, (IARC).
- Radmard AR (2010). Five common cancers in Iran. *Arch Iran Med*, **13**, 143-6.
- Robles SC, Marrett LD, Clarke EA, Risch HA (1988). An application of capture-recapture methods to the estimation of completeness of cancer registration. *J Clin Epidemiol*, **41**, 495-501.
- Sajadi A, Marjani H, Semnani SH, Nasseri-Moghaddam S (2010). Esophageal cancer in Iran: a review. *Middle East J*, **1**, 5-14.
- Schouten LJ, Straatman H, Kiemeny LA, Gimbrère CH, Verbeek AL (1994). The capture-recapture method for estimation of cancer registry completeness: a useful tool? *Int J Epidemiol*, **23**, 1111-6.
- Semnani S, Sadjadi A, Fahimi S, et al (2006). Declining incidence of esophageal cancer in the Turkmen plain, eastern part of the caspian littoral of Iran: a retrospective cancer surveillance. *Cancer Detect Prev*, **30**, 14-9.
- Shanmugaratnam K (1989). Availability and completeness of cancer registration worldwide. Recent results in cancer research. *Fortschritte der Krebsforschung. Recent Results Cancer Res*, **114**, 28-32.
- Suwanrungruang K, Sriplung H, Attasara P, et al (2011). Quality of case ascertainment in cancer registries: a proposal for a virtual three-source capture-recapture technique. *Asian Pac J Cancer Prev*, **12**, 173-8.
- Tilling K (2001). Capture-recapture methods: useful or misleading? *Int J Epidemiol*, **30**, 12-4.
- Tull ES, McCarty D (1992). Monitoring the incidence of myocardial infarctions: applications of capture-mark-recapture technology. *Int J Epidemiol*, **21**, 258-62.
- Wang X, Fan JC, Wang AR, et al (2013). Epidemiology of esophageal cancer in Yanting - regional report of a national screening programme in China. *Asian Pac J Cancer Prev*, **14**, 2429-32.